

SOFTWARE METAPAPER

MurCSS: A Tool for Standardized Evaluation of Decadal Hindcast Systems

Sebastian Iling¹, Christopher Kadow¹, Oliver Kunst¹ and Ulrich Cubasch¹

¹ Freie Universität Berlin, Berlin, Germany

MurCSS (Murphy-Epstein decomposition and Continuous Ranked Probability Skill Score) is a tool for standardized evaluation of decadal hindcast-prediction systems written in Python using CDO [1] and can be downloaded at <https://github.com/illing2005/murcss>. It analyzes decadal hindcast experiments in a deterministic and probabilistic way following and extending the framework suggested by Goddard et al. [2]. It was developed as part of the MiKlip (a major project for decadal climate prediction funded by BMBF in Germany) evaluation system to improve the comparability within the project during development stages and interim test phases. It is easily applicable by other modeling groups working on decadal prediction because it complies with international standards.

Keywords: Decadal prediction, skill score, MiKlip, MESS, CRPSS, model evaluation, verification

Funding Statement: *MurCSS* was developed during the research project MiKlip (FKZ: O1LP1160A) funded by the German Federal Ministry of Education and Research in Germany (BMBF).

(1) Overview

Introduction

In recent years decadal climate predictions have become more and more popular in the climate science community. Typically an ensemble of so-called hindcast experiments with starting dates between 1960 and today are performed enabling a verification against observational data. But the evaluation and validation of decadal climate prediction systems is both a scientific and a technical challenge in the current climate research.

There are several convincing arguments for using a standardized tool for decadal model evaluation. For instance, model development stages and interim test phases of (in this case decadal climate predictions) can be assessed easily. It also simplifies the comparison of decadal prediction systems developed by different modeling groups.

Goddard et al. [2] proposed a framework for verification of decadal hindcasts addressing two key questions:

1. Do the initial conditions in the hindcast lead to more accurate predictions of the climate, compared to uninitialized climate change projections?
2. Is the prediction model's ensemble spread an appropriate representation of forecast uncertainty on average

The first question is addressed through the Murphy-Epstein decomposition of the Mean Squared Error Skill Score (*MSESS*) [3],[4] which is based on the Mean Squared

Error and can be used to compare two different model outputs (e.g. initialized and uninitialized). The *MSESS* is a deterministic skill measure of the ensemble mean. For the second question Goddard et al. [2] suggest to use a modified version of the Continuous Ranked Probability Skill Score (*CRPSS*) [5], which compares the average ensemble spread with the mean squared error. We extended the probabilistic part with the logarithmic ensemble spread score [6], which is a direct estimate for the skill of the mean ensemble spread.

MurCSS follows this framework and offers a standardized and reproducible way to calculate the above mentioned metrics for decadal prediction systems including a bootstrap method to determine significance levels. A detailed explanation of the tool, its output, and a tutorial how to use the tool can be found on our web-page (<https://www-miklip.dkrz.de/about/murcss/>).

Implementation and architecture

MurCSS is a command line tool written in Python [7] using the Python-CDO-Interface [8]. The plots are produced and saved using the python library Matplotlib [9].

The tool architecture can be separated into three components (see **Fig. 1**), file input/preparation, metric calculation, and plotting routines. The file input searches for valid files and prepares them for the actual metric calculation. This component makes use of the MiKlip file database which is based on the international standards CMOR [10] and NetCDF [11]. It is also possible using the simplified file input component (`findFilesCustom.py`) which

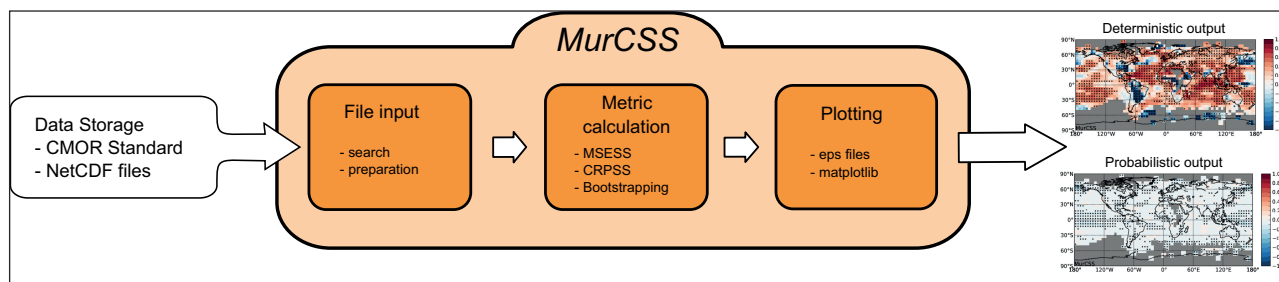


Fig. 1: The sketched work-flow of MurCSS with sample plots depicted on the right-hand side.

can be easily adjusted to the local database structure. The main part is the metric calculation, where the actual skill score calculation takes place. Most calculations are performed using CDO or the NumPy package, applying multiprocessing whenever possible. After processing the results are stored in the common NetCDF data format. In the last step these files are visualized using plotting routines.

Quality control

We developed the following strategy to assure the quality of the system.

- We assembled a set of test-data.
- Unit tests, which were constantly extended as soon as new problems were observed.
- Users within MiKlip helped as beta-testers to improve the tool.

(2) Availability

Operating system

MurCSS has been developed and tested on Linux. Although the tool should run on every system where Python and CDO are installed.

Programming language

Python 2.7, CDO 1.5.4

Dependencies

- Python Modules:
 4. Matplotlib >= 1.1.0
 5. Basemap
 6. NumPy >= 1.5.0
 7. SciPy >= 0.8.0
 8. CDOpy
- CDO
- NetCDF Libraries

Archive

Name

Figshare

Persistent identifier

<http://dx.doi.org/10.6084/m9.figshare.976060>

License

GNU General Public License 3

Publisher

Sebastian Illing

Date published

02/03/14

Code Repository

Name

GitHub

Identifier

<https://github.com/illing2005/murcss>

License

GNU General Public License 3

Date published

02/03/14

Language

English

(3) Reuse potential

MurCSS is already frequently used within the MiKlip project. The direct access to our file database and the standardized score calculation allows a quick assessment of the latest model improvements. The standardized output also facilitates the comparison of results obtained by different working groups within MiKlip. By now *MurCSS* has been used in a number of publications, for instance Pohlmann et al. [12] or Kadow et al. [13], and more in preparation.

Although it was developed to improve the comparability within the MiKlip project, we encourage other modeling groups working on decadal climate predictions to use our tool to benefit from the advantages that come with a standardized evaluation tool. In order to use the tool in projects outside of MiKlip *MurCSS* comes with a simplified file input component (`findFilesCustom.py`) which can be easily adjusted to the local data-structure in `murcss_config.py`.

The tool can also be applied to other disciplines of climate modeling, at the time of writing, for instance, we are working on an extension of *MurCSS* to seasonal forecast systems. Furthermore the modular architecture allows an easy extension to other skill scores or metrics of interest.

For the interested reader we established a guest account at our web-page (<https://www-miklip.dkrz.de/> / user: guest / password: *miklip*) where we provided some example analysis of the tool.

Support for MurCSS

When users or developers run into problems or discover bugs we encourage them to either open an issue on the GitHub page or to contact us via email (sebastian.illing@met.fu-berlin.de) directly.

We encourage users to submit pull requests on the GitHub page if they have developed new features.

Acknowledgements

We thank the two reviewers of this paper, who offered considerable constructive suggestions toward the improvement of this work.

We thank Madlen Fischer and Henning Rust whose comments helped to improve the tool and also the MiKlip-FLEXFORDEC (FKZ: 01LP1144A) project for providing the sample data of the MiKlip prediction system.

References

1. **Climate Data Operators** Max-Planck-Institute of Meteorology. Available at <https://code.zmaw.de/projects/cdo>.
2. **Goddard L, Kumar, A, Solomon, S, Smith, D, Boer, G, Gonzalez, P, Kharin, V, Merryfield, W, Deser, C, Mason, S J, Kirtman, B P, Msadek, R, Sutton, R, Hawkins, E, Fricker, T, Hegerl, G, Ferro, C A T, Stephenson, D B, Meehl, G A, Stockdale, T, Burgman, R, Greene, A M, Kushnir, Y, Newman, M, Carton, J, Fukumori, I, and Delworth, T** 2013, A Verification Framework for Interannual-to-decadal Predictions Experiments. *Climate Dynamics*, 40(1-2): 245-272. DOI: <http://dx.doi.org/10.1007/s00382-012-1481-2>.
3. **Murphy, A H** 1988 Skill Scores Based on the Mean Square Error and Their Relationships to the Correlation Coefficient. *Monthly Weather Review*, 116: 2417-2424. DOI: [http://dx.doi.org/10.1175/1520-0493\(1988\)116<2417:SSBOTM>2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(1988)116<2417:SSBOTM>2.0.CO;2).
4. **Murphy, A H and Epstein, E** 1989 Skill Scores and Correlation Coefficients in Model Verification. *Monthly Weather Review*, 117, 572-581. DOI: [http://dx.doi.org/10.1175/1520-0493\(1988\)116%3C2417:SSBOTM%3E2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(1988)116%3C2417:SSBOTM%3E2.0.CO;2).
5. **Gneiting, T and Raftery, A E** 2007 Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of American Statistical Association*, 102:477. DOI: <http://dx.doi.org/10.1198/016214506000001437>.
6. **Keller, J D, Hense, A, Kornbusch, L and Rhodin, A** 2010 On the Orthogonalization of Bred Vectors, *Wea. Forecasting*, 25: 1219–1234. DOI: <http://dx.doi.org/10.1175/2010WAF2222334.1>.
7. **Python** *Python Software Foundation*. Available at www.python.org.
8. **Mueller, R** *CDO{py,rb}*. Available at <https://github.com/Try2Code/cdo-bindings>.
9. **Hunter, J D** 2007 Matplotlib: A 2D Graphics Environment. *Computing In Science & Engineering*, 9: 90-95. DOI: <http://doi.ieeecomputersociety.org/10.1109/MCSE.2007.55>.
10. **Taylor, K E, Stouffer, R J and Meehl, G A** 2012, An Overview of CMIP5 and the Experiment Design. *Bull. Amer. Meteor. Soc.*, 93:485-498. DOI: <http://dx.doi.org/10.1175/BAMS-D-11-00094.1>.
11. **Network Common Data Format** *University Corporation for Atmospheric Research*. Available at <http://www.unidata.ucar.edu/software/netcdf/>.
12. **Pohlmann, H, Müller, W A, Kulkarni, K, Kameswarrao, M, Matei, D, Vamborg, F S E, Kadow, C, Illing, S, and Marotzke, J** 2013 Improved Forecast Skill in the Tropics in the New MiKlip Decadal Climate Predictions, *Geophys. Res. Lett.*, 40. DOI: <http://dx.doi.org/10.1002/2013GL058051>.
13. **Kadow, C, Illing, S, Kunst, O, Rust, H W, Pohlmann, H, Müller, W A and Cubasch, U** 2014 Evaluation of Forecasts by Accuracy and Spread in the MiKlip Decadal Climate Prediction System. *Meteorologische Zeitschrift, Special Issue on "Verification and process oriented validation of the MiKlip decadal prediction system"* (submitted).

How to cite this article: Iling, S, Kadow, C, Kunst, O and Cubasch, U 2014 MurCSS: A Tool for Standardized Evaluation of Decadal Hindcast Systems. *Journal of Open Research Software* 2:e24, DOI: <http://dx.doi.org/10.5334/jors.bf>

Published: 04 September 2014

Copyright: © 2014 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported License (CC-BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/3.0/>.

